DOCUMENT RESUME

ED 204 871 EA 013 800

AUTHOR McAfee, Jackson K.
TITLE Towards a Theory of Promotion: Does Retaining Students Really Work?
PUB DATE Apr 81
NOTE 30p.: Paper presented at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement: Age Grade Placement: *Educational Research: Elementary Secondary Education: *Grade Repetition: *Research Methodology: Standardized Tests: Statistical Analysis: *Student Promotion: *Test Interpretation: Validity
IDENTIFIERS Elementary Secondary Education Act Title I

ABSTRACT

Educators considering adoption of policies regarding the promotion of students from one grade to the next must consider several factors, including acceptable levels of student progress toward terminal objectives, student developmental and cognitive abilities, community attitudes toward student retention, and the educational benefits of retention. Research on the effects of grade retention has frequently been biased or at least inadequate, rarely using a true experimental design featuring random assignment of treatments to students performing at unacceptable levels. Misinterpretation of scores on standardized tests, due to skewed age-grade norms caused in part by variations in retention policies, has resulted in unreliable assessment of relative student progress. This attempt to assess student progress in terms of both age and grade norms used Model A, the frequently misunderstood evaluation design specified in Title I of the Elementary and Secondary Education Act. Analysis of the data showed that the retention policy under consideration was beneficial in the elementary grades, and had no significant effect on learning in junior high grades. High school grades were not tested. Statistical factors involved in analyzing the data are discussed, as are the needs for further research. (Author/PGD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED204871

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TOWARDS A THEORY OF PROMOTION: DOES RETAINING STUDENTS REALLY WORK?

JACKSON K. MCAFEE

SCHOOL DISTRICT OF INDIAN RIVER COUNTY

APRIL 1981

Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, California, April 13-17, 1981.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. K. McAfee
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EA 013 800

I. Introduction

Recent revelations of national studies on the decline in student achievement has aroused an interest in education and educational policies as the general public has become concerned over what appears to be happening to student performance in schools. Are tax dollars being efficiently spent? Are students graduating from school with undesirable levels of competency? Are students receiving high school diplomas that are functionally illiterate? These and many other questions are being asked as citizens worry and wonder about national declines in some areas of learning.

Legislatures have responded by enacting educational accountability statutes which prescribe procedures and means by which the efficiency and effectiveness of school instructional programs are to be examined and reported to the general public. Many states have enacted statewide assessment programs to provide quantifiable data on the extent to which students are achieving those skills or competencies that they are expected to achieve at various levels of instruction. As news of the alleged decline in the quality of education sinks into the consciousness of the citizenry, various lay groups are responding with intensified efforts directed towards revision of educational goals, curriculum philosophies and reliance on objective test data for making decisions on program adequacy, student promotion, and others.

As one of the issues, current interest centers on the philosophy of "social promotion", that is, the automatic advancement of children to the next level of instruction even though they may not have achieved all of the skills at the previous level. Citizens are demanding "performance based" systems of achievement that assures that students have attained certain levels of performance before promotion to the next level.

As various groups (educators, citizens, legislators, etc.) examine current goals and curriculum practices, the challenge becomes to avoid simplistic solutions which fail to address the underlying problems surrounding student achievement and the measure-

ment of that achievement. Educators realize anew the complexities involved in the learning process; the many factors both inside and outside the school environment which affect student achievement. Attempts to explore these complex problems in the hope of identifying professionally responsible solutions must face the wrath of lay groups who want simple answers and solutions and want them now.

The remainder of this paper will explore the impact of a particular promotion practice as it exists in a given school system, to explore underlying values and principles implicit in such practice and to gather empirical data on whether or not the practice was beneficial to students.

II. Organizing the Instructional Program

Instructional program designers must essentially deal with four components in designing instructional programs in given subject areas:

1. Identification of terminal student performance objectives that are consistent with the school system goals and philosphy.
2. Development of instructional activities that are designed to help students attain the terminal objectives.
3. Identification of resources (staff, facilities, materials) that are required to conduct the activities.
4. Evaluation directed towards determining whether or not students are attaining the desired outcomes.

It is assumed that citizen participation has been extensive in determining the goals and philosophy of the school system. Schools belong to the people. Americans pay their tax dollars to assure that students will receive instruction that perpetuates the American culture, that inculcates in students those values, beliefs, skills and competencies considered to be essential to maintaining our way of life. Schooling, thus conceived, represents a microcosm of what is happening in the larger society.

Education, as a public institution cannot expect to survive if it is not responsive to those ideals or goals which reflect the "mainstream" of American belief.

Given the identification of terminal objectives, it is recognized that each terminal objective may involve a number of skills that students are required to master before they have attained the terminal objective. Likewise, these skills generally exist in a hierarchy, with the mastery of some skills being prerequisite to the mastery of others. Below is an example of a terminal objective. Under the objective are given the skills which students must achieve in order to attain the objective. Also given are the levels of instruction at which we might expect each skill to have been mastered:

Terminal Objective - The student will divide whole numbers

Skill 1 The student will group twelve or fewer objects into sets of equal amounts (no remainder). (Grade 3)

Skill 2 The student will determine basic division combinations, dividends through 81. (Grade 4)

Skill 3 The student will divide a 3-digit number by a 1-digit number with remainder zero, without regrouping. (Grade 5)

Skill 4 The student will divide a 5-digit number by a 1-digit number. (Grade 6)

Skill 5 The student will divide a 3-digit number by a multiple of 10 less than 100. (Grade 7)

Skill 6 The student will divide a 5-digit number by a 2-digit number. (Grade 8)

The terminal objective represents a skill area which most Americans believe students should have attained by the time of graduation. These are skills deemed necessary by most to be functionally competent in adult life.

In organizing an instructional program for a particular terminal objective, the subskills must be identified, their heirarchy established (i.e., which skills are

prerequisite to which) and a sequence of presentation determined that establishes which skills will be taught at which levels of instruction that the terminal objective might be achieved by the time a student is the normal age for graduation. However, there is one other major consideration which must come into play in determining the levels of presentation for the skills, namely the students developmental level. Has the normal student matured to the point that he is capable of learning the skill at that point in time? Obviously, we do not attempt to teach calculus to fifth graders because most students have not yet developed the cognitive abilities which would allow them to master the skills associated with calculus. So it is with all subject matter and the process of determining the appropriate level of a given skill is nowhere near an easy one.

Historically the process of determining what is to be taught at a given level of instruction has been developed over a long period of time by the repeated experiences that teachers have had with teaching different skills at different ages. As a result of trial and error, an accumulated wisdom slowly specified the appropriate age for a given skill and instruction has been organized with this knowledge. Ironically, what has never been known is the extent to which all children of a given age are able to master the skills identified for a given level of instruction. It is suspected that there has always been a proportion of students that were unable to master the skills at that age though this varied from school to school and from classroom to classroom.

With the coming of age of educational research in this century and a commitment to the scientific method as a means of gaining knowledge about the learning process, much has been learned about the developmental levels of children and the ages at which the normal child can be expected to master certain skills. However, as of this date, there is still no widely recognized theory of learning nor is it apparent that what is known has been disseminated on a large scale so that it is common knowledge to educational practioners.

Given this state of affairs, the process of identifying the levels at which skills are to be taught is still primarily subjective and the use of empirical data on which one can systematically base decisions about what is appropriate has not been available. It is also recognized that a teacher's expectations about what are appropriate levels will be highly influenced by the caliber of students that a teacher has generally instructed. Thus, we can expect a third grade teacher in a high socio-economic suburban community to have different perceptions of what is appropriate compared to a third grade teacher in a low socio-economic rural community. Likewise, it can be seen that attempts to aggregate and average such perceptions on any population basis will automatically exclude a certain proportion of the population from being able to master the skills at the levels identified.

It is also apparent that given the distribution of academic ability that exists in a student population of a given age, there is a broad range of achievement between what the brightest students will attain and those who have severe mental limitations. To restrict one's curriculum to just those skills which all students of a given age should attain at a given level of instruction would result in a set of skills so minimal that most students would have already attained the skills before arriving at that level of instruction. At the same time progress must be made along a continuum of skills so that students will have attained the desired terminal objectives by the time of graduation.

Concomitant with the above, is the fact that teachers must also make decisions on individual students, to determine if they are ready for the next level of instruction. An appropriate promotion plan is to state those skills at given grade levels which are prerequisite skills that a student must have to perform sucessfully at the next grade level and to determine promotion on the extent to which those skills have been achieved. This would seem to be the most logical way for deciding whether or not a child should be promoted to the next grade. It must then be pointed out that not all of the skills that are taught at a given level of instruction are prerequisite

skills for performing successfully in the next grade nor should they be. If this were to be so, then enrichment activities would be excluded reducing the educational experiences that children could have at each level and restricting the educational opportunities to a set of minimal skills. It is doubtful that few would see this as a desirable alternative in an instructional program.

As educators struggle to identify appropriate skills at each level of instruction and to implement performance based systems of instruction, there are several alternative principles that have been used in the past. One is the following:

> The achievement of the skill is reasonable for every student in that grade in the basic education program who has appropriate instruction and motivation and is not impaired by a mental or physical handicap.

Strict application of that principle is apt to run into the following difficulties. In Florida, for example, only students who are in exceptional child programs are excluded from the basic educational program. Since by definition, entry of children into the educable mentally retarded program begins with an IQ of 69 or lower, it would be assumed by the above definition that children with an IQ of 70 or higher are capable of achieving the standards. Students whose IQ's fall in the range (70-90) are generally considered to be below normal in intelligence and experience great difficulty in learning basic skills at the same age level as their other peers. If we choose only those standards or skills for which it is felt that all students (IQ 70 or higher) could achieve, the resulting set of standards would be so restrictive or minimal that we could subject ourselves to a public mockery.

Another principle that has been used is the following:

> The achievement of the skill is reasonable for every student in that grade in the basic education program who has appropriate instruction and motivation and has normal or above intelligence (IQ 90 or higher).

Strict application of this principle can lead to the following difficulties. In the national population, roughly 23 percent of the students have below normal intelligence (within the range 70-90). Consequently, identification of skills on this basis could result in a retention rate of approximately 23 percent. This may be an intolerable rate in some communities.

Another principle that might be used which avoids the problem of skill identification is to say the following:

> On a national standardized achievement test, a student must attain a level of achievement greater than or equal to the national norm of that test.

This principal is appealing because of the inherent difficulties involved in attempting to define and measure specific skills at each level of instruction. A common way of trying to get around these difficulties has been to fall back on the comfortable fallacy that the grade norms that come with standardized tests can serve as appropriate promotion criteria. The criterion statement can be rephrased as follows:

"All fifth grade children should be able to read on the fifth grade level". Dyer states: "The fallacy in such a statement rests on two impossible assumptions; (1) that the fifth grade norm, that is, the average of all the scores that all fifth graders make on any reading test they happen to take - defines a level of reading performance that is to be regarded as in some sense satisfactory; and (2) that by some mathematical magic, the half of the student population which is below average can reach the average without raising it."[1] In short, if we are to promote pupils on the basis of whether or not their test score is greater than or equal to the national norm of that test, then by definition we are setting a performance standard which will automatically fail 50 percent of the students if the student population is similar to the national standardization sample.

III. Review of the Literature

The upshot of the previous discussion would seem to say that grade level achieve-

ment is whatever you define it to be. The three major considerations are (1) the sequencing of skills according to their heirarchy so that students will attain terminal objectives by the age of graduation; and (2) the developmental level of cognitive abilities of students at various levels of instruction; and (3) the setting of standards which will result in a percentage of retention that is acceptable to the community, that is, what failure rate the community is willing to live with.

Beyond that, are the implications of failure to the individual student and how this affects him in terms of self-concept (i.e., the affective dimension). Do the benefits gained by the child in increased acnievement as the result of retention outweigh the costs incurred to the child's sense of well being as the result of having failed? In fact, does increased achievement result from retention? One should be able to review research studies to determine what research evidence exists on the general topic of grade retention. Jackson has conducted a comprehensive review of the subject. "One general conclusion about the effects of grade retention relative to grade promotion is clearly warranted by all the results taken as a whole: There is no reliable body of evidence to indicate that grade retention is more beneficial than grade promotion for students with serious academic or adjustment difficulties".[2]

Thus, the past research on grade retention is inconclusive and it cannot be assumed that retention will provide greater benefits to students than will promotion to the next grade. Furthermore; Jackson states, "Thus, those educators who retain pupils in a grade do so without valid research evidence to indicate that such treatment will provide greater benefits to students with academic or adjustment difficulties than will promotion to the next grade."[3] Clearly, the need for research in this area is without question.

As Jackson discovered, much of the research on this subject has been quite inadequate for making valid inferences about the effects of grade retention. In fact, only

three studies were located that used the experimental design. "Investigators should try to arrange to use the experimental model in future research on this topic...those investigators who are inclined to use a nonexperimental approach to analyzing the relative effects of grade retention and grade promotion should proceed very cautiously, if at all."[4]

A review of the research since Jackson's article still reveals a lack of experimental research in this area.[5] The major difficulty in conducting an experimental study is that schools are unwilling to assign treatments randomly to pupils; that is, from a pool of potential retainees to randomly assign some to be promoted and some to be retained. This unwillingness results from the political realities that school people must face at the grassroots level. School officials are expected to know that retention is either effective or it is not. Given the lack of good data, one's position is partly determined by the most persuasive rhetoric that one hears. Schools must exhibit a degree of certainty in their decisions on student assignments. Parents, particularly, are not anxious to have their children as objects in a social experiment.

With the increasing move towards performance based education and the increased retention rates of students, it is imperative that empirical data be brought to bear on the question of whether or not retention is beneficial. And yet it is unlikely that an experimental design will be implemented by most school districts. Consequently, if evaluators are to reduce uncertainty regarding the policy of retention and its effects on students, efforts must be made to improve non-experimental designs so that threats to validity can be reduced and data can be brought to bear on the question.

IV. Methodology

Even a cursory examination of the studies of the effects of retention shows that most studies exhibit one of two biases. Either the design is biased towards showing that retention is more beneficial or towards showing that promotion is more beneficial. And to further complicate the matter, any school man worth his salt realizes very quickly that tremendous gains can be had in raising achievement test scores simply by the act of retaining significant percentages of students. By altering the age-grade distribution of students so that the mean age of each grade is increased with lower scoring students held back will almost inevitably lead to higher mean test scores for the grade.

Imagine then, the school principal who is faced with pressure from the superintendent, the school board, parent's advisory group, etc. What better way to raise test scores? This strategy does not call for any significant change in the instructional program; no brain-storming as to ways to improve the learning process in the classroom. All that is required is to simply hold students back, scores go up, and everyone is happy, except maybe some parents of retained students. But then, how can they complain, after all, achievement is up, the school is doing a better job, and furthermore, the principal has stated that students will not be promoted to the next grade until they have met a certain level of achievement in the previous grade. If one feels that this strategy is too farfetched to ever occur one need only consult Owen and Ranick[6] to find a true to life application of this strategy. Also, Hedges[7] refers to a similar application of this strategy.

Essentially, the problem that easily causes this misinterpretation of test results is discussed by Angoff.[8] For most national standardized achievement tests, publishers provide only grade norms as opposed to age and grade norms provided with aptitude tests. He states, "Because of the custom, prevalent in the early 1900s, of promoting children to a higher grade only if their achievement warranted, it was...

typical for the distribution of ages for children in the same grade to be highly dispersed and positively skewed. In order to standardize the population within grade and at the same time to make it more homogeneous, some test constructors suggested the development of modal age norms, that is to say, grade norms for children of approximately the same age...the effect of basing grade norms on such modal age groups was to free them to some extent of the influence of the local practices of retardation and acceleration; to move in the direction of greater homogeneity and precision."[9] Modal age norms have not been produced for some time, since the inception of "social promotion" made their use unnecessary. Nevertheless, if we are returning to an era where significant percentages of students will be retained it may be necessary to return to some form of age norms that we may properly evaluate educational growth.

The following provides one example of how a school system may analyze its test data using age groups as well as grade groups. Age group N represents the group of students who would normally be in grade N if they had not been retained or accelerated. Tables 1 and 2 give the results for two successive years of the age-grade distribution in relationship to Spring SRA Composite Scores (Mean Growth Scale Values) for a given school district. The results support the notion that the grade mean will generally be higher than the age mean for the age group normally associated with a given grade and that the grade mean is partly a function of the particular distribution of ages in the grade. Given this state of affairs, to accurately guage improvement from one year to the next, it would seem more appropriate to use age group means rather than grade means.

Such an approach as outlined above would represent a more honest way of reporting to the various publics on school district performance from year to year. However, given the political realities described previously, schools may be reluctant to report by age groups since the means are generally not as high.

Given the problems discussed above, how does one begin to resolve the question of whether or not retention is beneficial. First, we need to define when retention is

TABLE 1

1977-78 AGE - GRADE DISTRIBUTION IN RELATIONSHIP TO SPRING SRA COMPOSITE SCORES

AGE GROUP

GRADE GROUP		1		2		3		4		5		6		7		8		9		10				
		N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	
	1	584	124	149	111	8	115															741	121	1
	2	5	226	477	208	140	181	12	176													634	202	2
	3			6	277	482	260	127	217	12	199											627	250	3
	4					5	338	521	284	128	247	17	216	2	225							673	275	4
	5							5	291	480	330	114	281	8	248							607	319	5
	6									4	336	486	359	144	322	15	264	1	195			650	348	6
	7											5	314	527	401	155	344	9	295			696	386	7
	8													2	425	533	429	186	356	27	309	748	406	8
	9													1	585	4	462	529	464	183	384	717	444	9
AGE		589	125	632	186	635	241	665	269	624	310	622	340	684	382	707	407	725	434	210	374	6093		
GRADE		741	121	634	202	627	250	673	275	607	319	650	348	696	386	748	406	717	444					
		N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV			

14

	1	2	3	4	5	6	7	8	9	10

AGE GROUP

TABLE 2

1978-79 AGE - GRADE DISTRIBUTION IN RELATIONSHIP TO SPRING SRA COMPOSITE SCORES

AGE GROUP

GRADE GROUP	1		2		3		4		5		6		7		8		9		10			
	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV
	1	496	136	202	133	19	110														717	134
	2	5	255	455	219	181	191	28	190												669	210
	3			6	314	459	270	167	228	34	213	4	214								670	257
	4					7	327	438	306	188	252	40	232	3	230						676	289
	5					1	309	6	414	448	350	167	292	48	266	1	280				671	330
	6									7	413	442	396	135	308	37	278	3	275		624	370
	7											6	367	428	411	156	359	34	300	3	304	627	391
	8													3	342	514	445	224	368	64	328	805	414
	9													1	621	5	424	513	479	178	402	697	459
AGE		501	137	663	194	667	245	639	282	677	319	659	358	618	376	713	417	774	438	245	381		
GRADE		717	134	669	210	670	257	676	289	671	330	624	370	627	391	805	414	697	459				6155
		N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV	N	GSV		
		1		2		3		4		5		6		7		8		9		10			

AGE GROUP

considered to be beneficial. Retention is defined to be beneficial to the student if at the end of the year of retention the student showed an achievement gain relative to the group of students of which he previously was a member. If the student maintained the same or less relative standing, then the retention is defined to be detrimental.

A design that can help us answer the question is similar to the Model A evaluation design used in USDE Title I Evaluations.[10] Model A is a pretest-posttest design using either nationally normed or non-normed instruments. Where a non-normed instrucment is used, equating to a national norm-referenced test must occur in order that each test score can be expressed as a normal curve equivalent (NCE). Model A also requires that the same level of test be used both pre and post. Application of Model A results in looking for NCE gains to determine if the program has had an effect. Similarly, the present study can be designed to look for NCE gains from one year to the next to determine if retention has had a beneficial effect.

The use of Model A rests on the assumption that a cohort of students given no special educational intervention will, on the average, maintain the same percentile rank over the course of a school year as measured on a common standardized achievement test. This assumption allows one to estimate the no-treatment expectation as the percentile corresponding to the average pretest standard score. By comparing the actual posttest performance against the no-treatment expectation, one arrives at an estimate of the effect of the program. If the NCE gain from pre to post is positive, then the program has had an effect.

Hiscox,[11] Linn,[12] and Powell[13] have each in separate studies examined the validity of the assumption for estimating the no-treatment expectation. Their studies question the validity of this assumption based on selected samples they have studied. However, an elaboration of the logic involved in the assumption may clarify the issues; that is, the above studies may have been answering the wrong question. In the first place, the

assumption cannot be tested with data from most nationally standardized achievement tests. This is because most data based on fall and spring testing for the same grade is based on cross-sectional rather than longitudinal sampling of students. The effect of this becomes more acute in the upper grades where student attrition is greater (drop-outs, etc.). This may in fact explain why in some studies the mean percentile rank for some low socio-economic groups declines over time. With increasing selectivity in the population, the norms would be expected to show an increase.

It seems that a rationale for the assumption is based on the following. If we look at the entire student population for a given grade, the percentile rank of the posttest median will be the same as the percentile rank of the pre-test median by definition. This in turn establishes a rate of growth for the population at large in the grade. On the other hand, if we take a subgroup of students who are not progressing at the same rate of growth as the larger population, then with all other things being equal, we would expect this subgroup to maintain the same relative distance from the population mean on the posttest. Of course, in reality, all other things are never equal and for any particular subgroup we may select, they may actually gain more or less than the average gain for all cohorts in the country having the same characteristics as that subgroup.

The point is, that just as the difference in pre and post est medians defines an average rate of growth for the entire national group, so also does the difference in 20th percentile from pre to post define an average rate of growth for all students in the country scoring at the 20th percentile. If one of the stated reasons for the NCE metric is to provide a system for aggregating gains from Title I projects all over the country, then the assumption seems reasonable. That is, for some subgroups the gain will be greater than expected for others it will be less, but nationwide the aggregated gains should provide reasonable basis on which to judge the effectiveness of the Title I program.

If one is interested in comparing subdivisions of a local group to test effects of particular treatments, the assumption may indeed not be true. Essentially, when one calculates the NCE gain for a subgroup, one is comparing the growth of that particular group against the growth of all students in the country whose performance was similar on the pretest. A more acceptable interpretation of the Model A assumption then is that the pretest score establishes a baseline rate of growth against which the future performance of subgroups can be compared; the baseline rate of growth being the average rate of growth for all students in the country having the same pretest score.

Probably, the last thing one can say is that while there may be valid criticisms of Model A and the equi-percentile assumption, a cursory review does not reveal that anyone has posited superior models for consideration short of the experimental model. The fact that one cannot eliminate all of the threats to validity cannot excuse us from the central responsibility of having to evaluate the program. And evaluating the program is a must. Hence let us proceed with the rationale for Model A until such time as a superior non-experimental design can be developed which is practical to implement on a national scale.

V. Data Collection and Analysis

Application of Model A results in looking for NCE gains to determine if the program has had an effect. Similarly, the present study was designed to look for NCE gains from one year to the next to determine if retention had had a beneficial effect.

Data for this study were collected over a two year period (1977-78 and 1978-79) in a school district that had implemented a pupil promotion policy in grades 1-9 at the beginning of the 1977-78 school year and which resulted in 26 percent of the students being retained at the end of that year. Students were administered the SRA Assessment Survey both in April, 1978 and April, 1979. These two dates served as

the pretest and posttest. To comply with Model A requirements, data were collected at adjacent grades where the same level of the SRA test was administered both years. This resulted in collecting longitudinal data for grades 1 & 2, 4 & 5, 6 & 7 and 8 & 9. Students were separated into three groups for comparison purposes: The first group were those students who had been retained at the end of the first year. The second group were students in a compensatory education program and who were promoted after the first year. The third group were the remaining students who were promoted at the end of the first year.

Given in Table 3 are the mean NCE scores for each of the comparison groups. The mean NCE score was calculated as follows. For each of the groups, the mean growth scale value was calculated for the Composite Score on the SRA Assessment Survey. The total composite score is a weighted mean score obtained from the Total Reading, Total Language Arts, and Total Mathematics scores for each student. Once the mean growth scale value was determined, this was converted to a percentile rank using the norms tables provided by SRA.[14] In turn, percentile ranks were converted to NCE scores using conversion tables provided by USDE.[15]

An analysis of the data revealed the following. Retention appeared to have a beneficial effect in the elementary grades. In the middle-secondary grades no significant effect was observed. In all grades, students in compensatory education groups showed as large as or larger gains than the other groups. The remaining students showed gains in line with Model A expectations except for grades 2 and 7. The gain in grade 2 corresponds with the overall improvement in performance noted in that grade over the previous year. The decline noted in grade 7 reveals that this group did not gain as much as their national counterparts having similar scores on the pretest.

On limitation in interpreting the above results is as follows. Application of Model A implies that a selection test separate from the pretest has been used to place

Table 3

MEAN NCE GAINS FOR THE COMPOSITE SCORE ON THE SRA ASSESSMENT SURVEY (1977-78 and 1978-79)

		N COUNT	1977-78 MEAN NCE	1977-78 ADJUSTED MEAN NCE	1978-79 MEAN NCE	NCE GAIN
1st & 2nd GRADES	RETAINED	115	25	29	28	+3 (-1)*
	COMP. ED.	66	39	40	45	+6 (+5)
	REGULAR	343	63	63	67	+4
4th & 5th GRADES	RETAINED	84	25	29	30	+5 (+1)
	COMP. ED.	99	34	35	39	+5 (+4)
	REGULAR	287	62	62	62	0
6th & 7th GRADES	RETAINED	61	24	26	24	0 (-2)
	COMP. ED.	119	27	28	32	+5(+4)
	REGULAR	322	56	56	54	-2
8th & 9th GRADES	RETAINED	107	24	26	25	+1 (-1)
	COMP. ED.	79	28	30	31	+3 (+1)
	REGULAR	378	53	53	54	+1

*Based on the adjusted mean score

students in a compensatory education program. This is to control for the so called "regression effect" that results when low scoring students assigned to a program show a gain on the next testing whether instruction has occurred or not.

Another way to control regression effect bias is to statistically adjust the pretest means for the amount of regression expected based on the reliability of the test; the difference in the total and subgroup means; and the variance of the population and local group under consideration. This allows one to use the pretest as a selection test and to eliminate one round of testing. Roberts[16] has derived a formula that allows one to correct for regression bias as follows:

$$\bar{X}_{cs} = \bar{X}_g - (\bar{X}_g - \bar{X}_s)\left\{1 - \frac{\sigma^2}{s^2}(1 - \rho_{tt})\right\}$$

where

$\bar{X}_{cs}$ = corrected mean of subgroup

$\bar{X}_g$ = mean of total group

$\bar{X}_s$ = mean of subgroup (i.e., group selected for the program)

σ = population standard deviation

S = standard deviation of local group

ρ_{tt} = alternate form reliability

The use of this formula assumes that students have been assigned to the treatment based on their test score alone; that is, there is a cut off score on the test which divides the local group into two mutually exclusive subgroups. In the current study, analysis has been based on using the composite score on the SRA Assessment Survey. However, students were retained or promoted on the basis of other criteria as well; the composite score being only one of the criteria. Students were assigned to Compensatory Education Programs based on the SRA test, but the score used for selection was the Total Reading score in grades 1-6 and Total Language Arts or Total Mathematics scores in grades 7-9. Because placement in Compensatory Education or retention was based on multiple criteria, the regression effect bias would not be as severe as when decisions

are based on the Composite Score alone.

Table 4 demonstrates the degree of overlap between the promotion, retention, and placement of students and their composite scores. Particularly, in the lower grades, criteria other than test scores seemed also effective in determining the status of the students. Consequently, it is believed that regression is not operating to the extent that would be predicted by the formula. Nevertheless, the use of the formula does allow us to establish an upper bound on the extent of regression that has occurred and can establish a range in which we would expect to find the true means.

In Table 3 are given the adjusted mean NCE gains based on Robert's formula for correcting for regression bias. Adjusted means were calculated for both the retained and COMP ED groups. An analysis of these results reveals that the gains previously established for the retained students and COMP ED students are mitigated somewhat, more so for the retained students than for the COMP ED students. A further examination of Table 3, however, does not seem to warrant changing our initial conclusions but does serve to reduce somewhat the claim that retention is beneficial in the elementary grades.

TABLE 4

PERCENTAGES OF STUDENTS SCORING BELOW THE CUTOFF ON SRA COMPOSITE SCORE*
(1977-78)

	RETAINED STUDENTS	COMP ED STUDENTS
GRADE 1	79%	47%
4	85%	51%
6	84%	76%
8	81%	72%

*To meet one of the requirements for promotion, students were required to score at the fourth stanine or higher on the SRA Composite Score.

VI. DISCUSSION

The primary purpose and focus of this study was to begin a systematic investigation of what is considered one of the more pressing policy issues in education today. As more and more educational systems move in the direction of performance based instructional programs which cause significant percentages of students to be retained, it is important that we begin to answer the question of whether or not retention is beneficial to students.

It has become apparent to this author that the issue of social promotion is indeed one of immense complexity both in terms of how the instructional program is to be organized and how we measure the progress of students to effectively address the policy issue. With respect to organizing the instructional program, the determination of what are appropriate skills for students to master at each level has been primarily a subjective one In that regard, the process of assessing students has been one of objectively determining the extent to which children have achieved those things which we have subjectively determined they should. If anything, the process of developing a performance based system should reveal to the astute how little we really know or understand about how children learn.

Be that as it may, our lack of knowledge on which to base responsible decisions has not prevented school systems from forging ahead with performance based systems. Given this state of affairs, we must nevertheless attempt to answer the question of how effective these systems are. It is recognized at once that the use of traditional national norm-referenced achievement tests with their reliance on grade norms does not allow us to answer the question very well. When significant percentages of students are retained, the age-grade distribution of students is altered and confounds immensely the interpretation of one's results against grade norms. It seems necessary that test publishers begin developing age norms for achievement tests such as they have done for aptitude tests if we are to accurately gauge the progress of students across time. At

the least, school systems can collect data on age groups such as presented in this paper as one means of improving the interpretability of their results.

To determine whether or not retention is beneficial, all would agree that implementation of experimental designs would best allow us to answer the question. Unfortunaly, it seems that most school districts will be unwilling to adopt such a strategy because of the political ramifications. If this is so, then efforts must be made to improve non-experimental designs so that threats to validity can be reduced. Such a design has been presented in this paper. There is no doubt in this author's mind that designs superior to this one can be developed and would encourage researchers to improve on the present design. In any event, it is important that the work of systematically collecting empirical data on the effects of retention begin immediately. If in fact, we have embarked on an ill-conceived educational policy and no attempt is made to adequately evaluate such policy, then the social cost to retained students will be irredeemable not to mention the tax dollars lost by keeping those students in school beyond the normal number of years.

The results of the current study suggest that retention appears to be beneficial only in the elementary grades. However, given the nature of the non-experimental study, there are caveats that limit the degree of certainty one can attach to the results. Hopefully, if designs such as this one (or improved designs) could be implemented on a broad national scale, then a meta-analysis of the results could suggest directions for future policy action.[17] It seems imperative that we embark on such an endeavor.

As ways of improving the present design are considered much attention should be focused on the emerging RASCH technology for test construction.[18] Such problems as cross-sectional data on which the norms are based and the effect of student attrition; floor effects; out-of level testing effects; poor fit between the curriculum and the test; and scaling problems may more effectively be dealt with through the use of the Rasch Model. In fact, the use of the Rasch Model may enable us to empirically solve the more

important problem of identifying appropriate skills at each level of instruction.

While the present study lacks the rigor normally associated with an experimental design; a design which has been used nationwide was employed. With the increasing move towards performance based education and the increased retention rates of students, it is imperative that empirical data be brought to bear on the question of whether or not retention is beneficial. Consequently, evaluation studies, such as indicated here, provide one tool for evaluators to use in helping to reduce uncertainty regarding the policy of retention and its effects on students.

Footnotes

1. Henry Dyer, How to Achieve Accountability in the Public Schools, Phi Delta Kappa Educational Foundation - Fastback No. 14, 1973, Bloomington, Indiana.

2. Gregg B. Jackson, "The Research Evidence on The Effects of Grade Retention", Review of Educational Research, Fall 1975, Vol. 45, No. 4, p. 627.

3. Ibid.

4. Ibid., p. 628

5. Student Retention vs. Social Promotion: The Best of ERIC, No. 43, 1979. (ED 166 771).

6. Samuel A. Owen and Deborah L. Ranick, "The Greensville Program: A Commonsense Approach to Basics", Phi Delta Kappan, Vol. 58, No. 7, March, 1977.

7. William D. Hedges, "How to Raise Your School's Ranking by 172 Places in One Year Without Changing the Curriculum or Antagonizing Your Teachers", Phi Delta Kappan, Vol. 60,No. 5,January, 1979.

8. William H. Angoff, "Scales, Norms, and Equivalent Scores", Educational Measurement Second Edition, Edited by Robert L. Thorndike, American Council on Education, Washington, D.C., 1971.

9. Ibid., p. 538

10. G.K. Tallmadge and C.T. Wood, User's Guide, ESEA Title I Evaluation and Reporting System , RMC Research Corporation, 1976.

11. S.B. Hiscox and T.R. Owen, "Behind the Basic Assumption of Model A", Paper presented at AERA/NCME Annual Meeting, Toronto, Canada, March, 1978. (ED 166 258)

12. R.L. Linn, "The Validity of Inferences Based on the Proposed Title I Evaluation Models," Paper presented at AERA/NCME Annual Meeting, Toronto, Canada, March, 1978. (ED 156 696)

13. George Powell, Jerome Schmidt, and Paul Raffeld, "The Equipercentile Assumption as a Pseudo-Control Group Estimate of Gain", Paper presented at AERA Annual Meeting, San Francisco, California, April, 1979.

14. SRA Assessment Survey: Rights Keys, Norms, and Conversion Tables - Primary and Multilevel Editions, Science Research Associates, Inc., Chicago, Ill., 1974.

15. Tallmadge, op. cit.

16. A.O.H. Roberts, Regression to the Mean and the Regression Effect Bias, RMC Research Corporation, Mountain View, California, 1980.

17. For an example of how a meta-analysis could take place refer to a seminal article by Gene V. Glass, "Primary, Secondary, and Meta-Analysis of Research", Educational Researcher, November, 1976.

18. An excellent primer on the subject of the Rasch Model is by Benjamin D. Wright and Mark H. Stone, Best Test Design, MESA Press, Chicago, Ill., 1979.

19. For an example of an empirical skill continuum, refer to, Skill Continuums in Mathematical Computation, Reading Comprehension, and Written Expression, Los Angeles County Test Development Center, Office of the Los Angeles County Superintendent of Schools, 1979.